Lab Report

1. Optimal Clusters:

The optimal number of clusters we identified from the elbow method is 5 as the "elbow" point was at k=5. by the hierarchical clustering, i.e from the dendogram we observed that the optimal clusters seems to be 3 however we see that more information would be lost if we cut a horizontal line that high as there are long jumps beneath it and not considering them would be a mistake, hence we observe that 5 clusters are the optimal number of clusters.

2. Cluster Comparison:

Figure 1: K-means

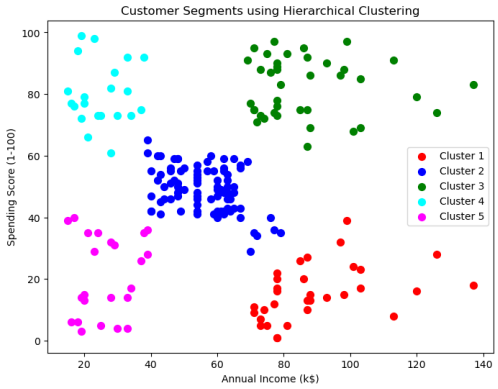

Figure 2: Hierarchical

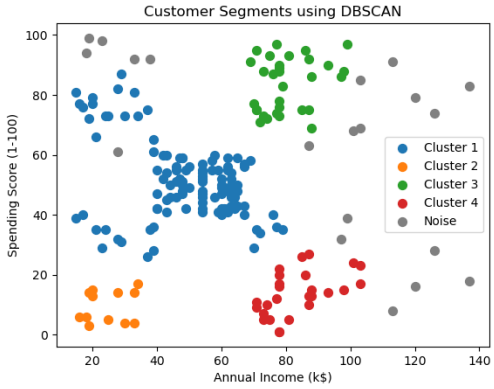

Figure 3: DBSCAN

K-Means and Hierarchical clustering produced very similar intersections/clustering, both of the methods separated the customers into clear income–spending groups. DBSCAN, with parameters eps=10, min samples=6, gave less number of clusters and also identified noise from the data. Unlike K-Means/Hierarchical Clustering, DBSCAN did not include each data point into a cluster.

3. DBSCAN performance:

DBSCAN identified and segemented the data into density based clusters using the paramters: eps = 10 and min sample = 10. It also identified/isolated the noise in the data by taking noise as -1 and bifurcated the data into 4 clusters. As compared to K-means and hierarchical clusering, DBSCAN gave a non-spehrical clustering pattern.

4. Algorithm Suitability for this Dataset:

The k means and the hierarchical clustering methods provided with a spherical clustering pattern with well shaped clusters. As the DBSCAN is for the data with heavy noise and irregularity, provided with comparatively fewer clusters and isolated/categorised some data points as noise too. Overall, for the customer data of the dataset "Mall Customer Segmentation" the K-means and hierarchical methods are more suitable as they provide with the proper classification of the clusters and this data isn't very noise heavy.

5. Real-World Application:

For particularly this dataset the mall's marketing team can use the clustering data information to design various schemes and campaigns to engage more customers. The biggest cluster shows a particular customer base which belongs to the same category of shoppers and hence can design more lucrative schemes for them to not churn and for the rest, according schemes can be made which can attract more customers, This analysis can help the business run in a more efficient manner.